

Abstract

- 69 -

0000572.050499

| | | | | |
|---|-----|-----|-----|------|
| | 150 | 155 | 160 | |
| GCG ATG GTA ATG GAT CGG CTA TAC GGA GGT GTG TGC TAC GCT GGG ATT | | | | 586 |
| Ala Met Val Met Asp Arg Leu Tyr Gly Gly Val Cys Tyr Ala Gly Ile | | | | |
| | 165 | 170 | 175 | |
| GAT ACC GAC CCT GAG CTA AAA TAC CCA AAA GGA GCT GGG AGA GTT GCG | | | | 634 |
| Asp Thr Asp Pro Glu Leu Lys Tyr Pro Lys Gly Ala Gly Arg Val Ala | | | | |
| | 180 | 185 | 190 | |
| TTC TCT AAT CAA CAG AGT TAC ATA GCT GCT ATC AGT GCC CGC TTT GTT | | | | 682 |
| Phe Ser Asn Gln Gln Ser Tyr Ile Ala Ala Ile Ser Ala Arg Phe Val | | | | |
| | 195 | 200 | 205 | 210 |
| CAG CTG CAG CAT GGA GAG ATA GAT AAA CGG GTA AGC CTT ATA CTA CAT | | | | 730 |
| Gln Leu Gln His Gly Glu Ile Asp Lys Arg Val Ser Leu Ile Leu His | | | | |
| | 215 | 220 | 225 | |
| TTT GGA AAA TTC TAGAAATGGT CCTCTAAATG TGTGATTACC AATATTAGAA | | | | 782 |
| Phe Gly Lys Phe | | | | |
| | 230 | | | |
| CGGGAGCATT TTATGACAAT AAAGTGACAG CTGACAATTT TGCCTATAGA GTTAATTATG | | | | 842 |
| GTCTATAATA CATGAAATAA TGTCTATGA ATTTCTTTTA TCTTTCAGTT TTTTGAGTAG | | | | 902 |
| CCTAATCAGA AACTACAAT TACTTGAGT TAATTTAATC TTCTCTAACT TCCATTCAAT | | | | 962 |
| CTCAATCCAT CCGTCCATT ATTCACTTAG TTGTAAGTC ATTCAATAAA TATTTACTGA | | | | 1022 |
| ATCCTTTTGT CTGTGTTATA TCAAGTATAC AAACAGGAAT GCCCTTGAGG TTCTCTGCC | | | | 1082 |
| TTTTTTTTGT TTGTTTTTTA ATCTGGGAC ATAGGGAAGA CCTCAGCAAG CCTATTCT | | | | 1142 |
| CAATGAATTG TACTCACAGA TTCTTTTTTT TTTTTTTTTT TCTTTTTCCA CAGCGCCAC | | | | 1202 |
| CTCTCACC GA TTTATTCCTT AGCTTGGTGT TTCATGTATT CAACAAACGT TTAGTGCTT | | | | 1262 |
| AGGGCAAGAA GTTCTGTCC TCATGAGTTT ATTTCTAGC AGATAGAACT GTATCACTTG | | | | 1322 |
| CCAGTACTAC TCAGAGTGTG GCCTGTGGAC TGACCTCCAG TCTGTAACT TAGTTGTAG | | | | 1382 |
| TGAGATAGGA ATTTAGACCA GAATGTGTAA TCAACCACAT TACTGGGCAC AATGTTTGGT | | | | 1442 |
| CCAGCTGGCG ATTTTTTTTT CATAGAAAGC CTTTATTGAT GAGGGAAGCA ATATATTGAT | | | | 1502 |
| TTATATTTTG GGGTCACTT TTTATTTTAT GGCACACTGG CACTTTCATG CATGCTGACT | | | | 1562 |
| TTGATATCCA TCACTCTGAG GCATTGTGCT AAAATAGATT GATTTTATCG TGTGTTCTC | | | | 1622 |
| AATTCAAGAT GTAAAAATCA TCAAGTCAGT AGCAGTTTTT GCTTTTTATG TTTTATGTCA | | | | 1682 |
| TGTACAGTCT ACTTCACTGG CAGTAAAAAA ATTTAAGATA GTGGTGGTCA TCCTACAAAC | | | | 1742 |
| TGTGAATCTA TTAAGAGAA AAGTATCTGT TCTATTCTAA GCATGGGGGA GGGACAAGAT | | | | 1802 |
| TAGTATGTTA ACATGCCTAC TTGTTTTGTT TGAGATGGAG TCTCTCTCCG TCACCCAGGC | | | | 1862 |
| TGGAGTGCAG TGGTACAGTC TCAGCTCACT CCAACCTCTG CCTCCCGGGT TCAAGTGATT | | | | 1922 |
| CTCCTGCCTT AGCCTCCCGA GTAGGTGGAA TTACAGGCAT ATACCACCAT GCCCAACAAA | | | | 1982 |
| TGTTTGTATT TTTAGTGAG ACAGGGTTTC ACCGTGTTGG TCAGGCCAGT TTCAAACTCC | | | | 2042 |
| TGACCTCAAG GGATCCACCT GCCTCACCCC CTCAAAGTGC TGGGATTACA GGCATGAGCC | | | | 2102 |
| ACCCACCATG CCTGGCCTAC TTGTTTTTTT ATGCACACTA AAAAATACCT ACATCTCACT | | | | 2162 |
| GCCTTATTCC AACATAAGTT TCAGAGCTGT GGGATTGGTC ATTAGAAAT CAGACTGAAT | | | | 2222 |
| TTGTGTTCT CTGCAATGAA ATCCTTTGCC CAGTGTTCAT GTCACCTGT AGACATTATG | | | | 2282 |
| GAGCAGCCTA GAGGCCAGAA GCCCAGTGT CTCCTTATGC CTGCTCTTCC TGGGCTTCGT | | | | 2342 |
| GACACTCTTC TTCTCCTTTT GTACTTTTAT TTTTATAGTT AAAAAATTTT TTTTAGAGGG | | | | 2402 |
| AGGGTCTCAC TCTGTCACCC AGGCTGGAGC ACAGAATCAC AATCATGACT CACTGCATGT | | | | 2462 |

| | | | | | | |
|--------------|--------------|--------------|-------------|--------------|-------------|------|
| TTCTCTCCTT | TGTGTCATGG | CTAATCTTGG | TCAGGATTCC | TTGTCAGAGC | TGGGTGGCAC | 2522 |
| CAGTGTCTGGT | GACAGCCTGC | TGTAAGGGAG | TTTCAGCCAT | GAATCTCTCC | AGACTAAAAA | 2582 |
| TAACCAGCTC | TTTTCTAGCT | GATGAATTAA | TAACCAGGTG | ACTGTTAATG | CTTGAAAGGT | 2642 |
| TCACATGACA | GGTTGGCCGA | TAGAACGCTG | GAACAGGCCC | AGTTTATGAA | ATTCACCTCT | 2702 |
| GACTTTTAGA | CTCAGGTGAA | CCATTCTTAC | TGAGAAAGAA | CAAAAGCAGG | TTTTAGACTG | 2762 |
| TGAATCCAT | GGCTGCATCT | TTTTTTTTTT | TTTAACAGAG | TTCAGGTTT | GTGATTATAA | 2822 |
| CCCCAACATG | GTACACTATA | AATAGAAACC | ACGAGCCAGG | CTTTTACGA | CAGCTCAGAA | 2882 |
| TTCTGTGAGC | GAGTAGTCAG | GCATCTTAC | ACCGACTTGA | ATATTGAAGT | GCGAGTGTGT | 2942 |
| GGAAC TTGGA | TCATCTTAGT | TGATTTTGTT | TAAATTATGA | TCCACATAT | GACAAAAATC | 3002 |
| CAGATCCACT | AATTA AAAATG | AGGGTTTATG | TCTATGAATA | ATCTCCTGTG | GGTTTAATCT | 3062 |
| CATAACATTC | TAGTCTAAAC | AGTTGGCTTC | ACTTCATGAT | GTCGTGCTAA | ATCCTTTTTC | 3122 |
| CTTTAAAGGA | TGTTTTATTA | ATAAGAAAAA | AAATGTAAAA | TGATAGATAA | TAAAAAGCCT | 3182 |
| ACTAGGTTCT | TAAAAGATGA | ACTATCCATA | TTTCAGTAA | TGAATAATTA | GTCCTTCTCT | 3242 |
| TTTGGGCACC | TGGGAACAGA | TTCATTCAGA | TAGTGGGTGG | AAATGTACAT | GTATGGTAAG | 3302 |
| CAGTCTGGC | CTAGTCACT | AAAAATGTAA | ACTCTTATTT | TGATTTGACG | GTGGAAGTTA | 3362 |
| AGCCATATGT | CTTGATGTAT | CAGCTGTGTG | ATGAATGTCA | GGGGGCCCGT | TGTGGGGGGA | 3422 |
| AATTTGCTCC | ATTTTTCTGT | GCTAATGTTA | CCTGCTCGCA | GTATTACTGT | GAATATTGCT | 3482 |
| GGGCTGCTAT | CCATTCTCGT | GCTGGCAGGG | AATTCACAA | GCCCCTGGTG | AAGGAAGCGG | 3542 |
| GTGACCGCCC | TGGGCATATT | TCATTCGCC | GGAAC TAAAG | GATAACTGCA | GTGCTCATTT | 3602 |
| TCAGGCCCTCA | GAATAAGTGC | ACTCTTCTGT | TCATTCTGAC | CCCTTCTCTA | ACCTCTTCAC | 3662 |
| GCTGGCATGT | CCTTTTGTAG | CAGTCTGTGA | CTTAAC TATA | G TATAATGAA | AAGAATGACC | 3722 |
| TATAATATAG | TGTTTTGTGA | GATTTCTGTG | TCTGTCAAA | CAATATGAA | TCCTTTTTTCG | 3782 |
| TATTGCCATC | GGGTGTCATG | GAGGTTTTAT | TCCTTGTGTT | TGCTGGAAC | CACAGGATC | 3842 |
| CAAACTTCTCT | GCAACATTTT | CTAGAGGAG | AGAGAGAAAT | ATTA AAAAGAG | AAATGAAACA | 3902 |
| ATAGAGTATT | TTGGGTTTTT | AATTA AAATTA | TTGTTAATAA | TATAACATAT | AAGAATACTT | 3962 |
| TTATTA AAAAT | AACCATGCAA | CAATAACACT | ATCGGTCTAT | CTGACAGTTT | TCCCCCAGG | 4022 |
| GAAGTGCTTT | TGCTTTTTTC | TTCTTTTTTT | TTTTTTTTTC | ATCTTTTTTG | TCTCTCTCT | 4082 |
| TTTTTCCATC | CCTTTTTTAAT | TTTTTTAACA | GCAATGGAGG | AAGTTAACAA | TTTTTAATGG | 4142 |
| AAAGAGCATC | TTAGAGCAAA | CAAAATGCATA | AGCAAGACTG | AGCAGCATTA | TAATTAATTT | 4202 |
| TCAGGGTTTT | GAGGCTGAAC | ATAATTTTCAT | TATCCCTCAA | AAAGTTACCA | CCACATCAGA | 4262 |
| AAAAAA AAAAA | AAAA | | | | | 4276 |

- 71 -

| Met Gly Ser | | | | | | | | | | | | | | | | |
|-------------|-------------|------------|------------|------------|------------|------|-----|-----|-----|-----|-----|------------|-----|-----|-----|-----|
| 1 | | | | | | | | | | | | | | | | |
| GAC | AAA | AGA | GTG | AGT | AGA | ACA | GAG | CGT | AGT | GGA | AGA | TAC | GGT | TCC | ATC | 163 |
| Asp | Lys | Arg | Val | Ser | Arg | Thr | Glu | Arg | Ser | Gly | Arg | Tyr | Gly | Ser | Ile | |
| 5 | | | 10 | | | 15 | | | | | | | | | | |
| ATA | GAC | AGG | GAT | GAC | CGT | GAT | GAG | CGT | GAA | TCC | CGA | AGC | AGG | CGG | AGG | 211 |
| Ile | Asp | Arg | Asp | Asp | Arg | Asp | Glu | Arg | Glu | Ser | Arg | Ser | Arg | Arg | Arg | |
| 20 | | | 25 | | | 30 | | | 35 | | | | | | | |
| GAC | TCA | GAT | TAC | AAA | AGA | TCT | AGT | GAT | GAT | CGG | AGG | GGT | GAT | AGA | TAT | 259 |
| Asp | Ser | Asp | Tyr | Lys | Arg | Ser | Ser | Asp | Asp | Arg | Arg | Gly | Asp | Arg | Tyr | |
| 40 | | | | 45 | | | | 50 | | | | | | | | |
| GAT | GAC | TAC | CGA | GAC | TAT | GAC | AGT | CCA | GAG | AGA | GAG | CGT | GAA | AGA | AGG | 307 |
| Asp | Asp | Tyr | Arg | Asp | Tyr | Asp | Ser | Pro | Glu | Arg | Glu | Arg | Glu | Arg | Arg | |
| 55 | | | 60 | | | 65 | | | | | | | | | | |
| AAC | AGT | GAC | CGA | TCC | GAA | GAT | GGC | TAC | CAT | TCA | GAT | GGT | GAC | TAT | GGT | 355 |
| Asn | Ser | Asp | Arg | Ser | Glu | Asp | Gly | Tyr | His | Ser | Asp | Gly | Asp | Tyr | Gly | |
| 70 | | | 75 | | | 80 | | | | | | | | | | |
| GAG | CAC | GAC | TAT | AGG | CAT | GAC | ATC | AGT | GAC | GAG | AGG | GAG | AGC | AAG | ACC | 403 |
| Glu | His | Asp | Tyr | Arg | His | Asp | Ile | Ser | Asp | Glu | Arg | Glu | Ser | Lys | Thr | |
| 85 | | | 90 | | | 95 | | | | | | | | | | |
| ATC | ATG | CTG | CGC | GGC | CTT | CCC | ATC | ACC | ATC | ACA | GAG | AGC | GAT | ATT | CGA | 451 |
| Ile | Met | Leu | Arg | Gly | Leu | Pro | Ile | Thr | Ile | Thr | Glu | Ser | Asp | Ile | Arg | |
| 100 | | | 105 | | | 110 | | | 115 | | | | | | | |
| GAA | ATG | ATG | GAG | TCC | TTC | GAA | GGC | CCT | CAG | CCT | CGC | GAT | GTG | AGG | CTG | 499 |
| Glu | Met | Met | Glu | Ser | Phe | Glu | Gly | Pro | Gln | Pro | Ala | Asp | Val | Arg | Leu | |
| 120 | | | | 125 | | | | 130 | | | | | | | | |
| ATG | AAG | AGG | AAA | ACA | GGT | GAG | AGC | TTG | CTT | AGT | TCC | TGATATTATT | 545 | | | |
| Met | Lys | Arg | Lys | Thr | Gly | Glu | Ser | Leu | Leu | Ser | Ser | | | | | |
| 135 | | | 140 | | | | | | | | | | | | | |
| GTTCTCTTCC | CCATCCCCAC | CTCAGTCCCT | AAAGAACATC | CTGATTCCCC | CAGTCTTCAA | 605 | | | | | | | | | | |
| GCACATGAAT | TCAGAAATGAA | AGGTTTGCCA | TGGCTAAGGA | ATGTGACTCT | TTGAAAACCA | 665 | | | | | | | | | | |
| TGTTAGCATC | TGAGGAACCTT | TTTTAAACTT | TGTTTTAGGG | ACTTTTTTTT | CCTTAGGTAA | 725 | | | | | | | | | | |
| GTAATGATTT | ATAAACTCCT | TTTTTTTTTT | TTGACTATAG | TCGGTTGCAT | GGTTACTTTA | 785 | | | | | | | | | | |
| AGCGTGGAAT | CAATGGAGT | GGCATTTAGT | TCAGGCGGGT | TGTTCCCTTG | CATGGCAAAG | 845 | | | | | | | | | | |
| TATCAAGAAG | ATCCCCAAGT | CAAGTCACAT | TTGTAAGATC | GCTTCCCAAT | TGGCTTTGTC | 905 | | | | | | | | | | |
| ACGCAGTGTT | GAAGCAGTGG | GAGAGAGATT | CACCTGTAT | AAGGAACTG | ACTAACACAA | 965 | | | | | | | | | | |
| GTATCCCGTC | TATATCTGAA | TGCTGTCTCT | AGGTGTAAGC | CGTGTTTCG | CCTTCGTGGA | 1025 | | | | | | | | | | |
| GTTTTATCAC | TTGCAAGATG | CTACCAGCTG | GATGGAAGCC | AATCAGGTTG | CTTCACTCAC | 1085 | | | | | | | | | | |
| CAAGTCTAGA | TATTCATGAA | AATGGAACAA | GTCTGTACAA | TTTTAAAAAA | AGGTTGAAGG | 1145 | | | | | | | | | | |
| AGTGGTTTGT | TCCAAAGGAG | TGACTTTTTT | TTAAAAAAA | AAGCTTTGTA | TATATTAAAA | 1205 | | | | | | | | | | |
| TIGATGTTAC | TAGAATAAGT | ACAGTACCAA | GGACTTCATT | ATAGAATTTG | TTCTGCGCTT | 1265 | | | | | | | | | | |
| AAACATGGCT | ACCTACCTGG | | | | | | | | | | | | | | | |

```

GGCGGCATT TTATAAGTAA TTCCTTATTT CTGCCTGAAG ACCACAAAGC CTCCTGGAGG 1505
CGTAACTGCT CAGACCGGTC TTCAGGGAAT ATTTAAGGAC TTAGTGGAAT TTATGAACAA 1565
TAAGTCTGAT GAGATTAGCC TGGGAGTGGT GTCCTGCAGC TGCTAATCT AGTTAGAGTG 1625
GCATTACAT TCTAATCTCC TTGAGAAATG CTTTATAGT CTGTTCAAAG CAAGTCATTG 1685
ATGGTTCTTC GAGGTAGTGT TAACTGAAGT GTTCTTCAGT TTGTCAAGAT AATGTTTCAGT 1745
GCTTGGCACT TAAATAACAT TTTTGCAGG AACTCCAAGG CACATTATTG AATGCCTTTA 1805
ACCAAGTGCA TTCTGGGAAG TTTGCTTGAC TCATTATCTT GCTTTTCTGC AGCATTCTGT 1865
GATTGAGTC ATCCATGAAT CCATGAATAA AAGTTACATT CTTTGATTGG TAATATTGCC 1925
ATTATAACA AGACTCACTA ATGAGGGTAT CACTTTGACT GACTGATTGG TTAAGTTTTT 1985
TAAGCCTCTC ATTTTCCTAA CCCAGAAATC ACAGCCTGAT TTTATTAATA GTAGAGCTTC 2045
ATTCAATTTA TACCATAGAT ACCATCCTAG TAAATCCAGA ACATATACAA GGTTCATGTG 2105
AGTCTGCTTT CTTGACATGA TAGCAATTGT TGATGCAGTG GATATGTCAG AATGACTAAC 2165
CTAGGAGTTT AAAACTCCTA AGAACTAAA ACCTGAAGA CATTTAAAAG TCTCCACAA 2225
TTTAATGTAT ACAAAGCTAT GTTACTGTGT AACACATTAC AGTTCAAATT CACTCCAGAA 2285
ATAAAAGGCC AGTAGGATTA GGGACTCACT GGTAGTTTGG AGTCTCCCAG CACACATCCC 2345
TCCTAGTGGG ATGATCTATT CACATATCTC CCAGCTTTTT TATTTTGTCT TCTGTATATC 2405
ACAGTGAGTG GATGGCCCTT CAGCTTTTTT TCTCCTGGCC AGACATGCAG TCTTGCCCTT 2465
AGATATCGCA GAGACAAAAT TCACAGCATG TCTTAAATCT TCCAGGATTG GCAAGAACCA 2525
AATTGCTCAA CAGTATGTAT GTTTAGAGGG GTTAGACTCC TTTTAAAAAT CTGGATATCT 2585
AACCACCTAC TTAATCTGTG TTGATAGTGT CAAACCACCC CCACCCCTGA TCCTCCACCC 2645
CCCAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAA 2689

```

SEQ ID NO:3
 SEQUENCE LENGTH: 2981
 SEQUENCE TYPE: nucleic acid
 STRANDEDNESS: double
 TOPOLOGY: linear
 MOLECULE TYPE: cDNA
 ORIGINAL SOURCE
 ORGANISM: human
 CELL TYPE: leukocyte
 SEQUENCE DESCRIPTION: SEQ ID NO:3

```

CCTCTCTCTC TCTTTCACAG AGTCTTGCTC TGTCGCCAG GCTGGAGTGC AGTGGCACAA 60
TCTCACTGCA AGCTCCGCCT CCGGGTTCA CGCCATTCTC CTGCTCAGC CTCCCAAATA 120
GCTGGGACAA CAGGCACCTG CCACACGCC CGGCTAATTT TTTGTATTTT TAGTAGAGAC 180
AGGGTTTCAC CATGTTAGCC AGGATGGTCT CAATCTCCTG ACCTCGTGAT CCACCGCCT 240
CAGCCTCCCA AAGTGCTGAG ATTACAGGTG TGAGCCACCA CGCCAGCCA CATCTTCTT 300
TCTTTCTTTT TGGTTTTTGT TTGTTGTTG AGACAGGGTC TTGCTCTGTC GCCCTGGCTC 360
ACGTGAACCT CCCACCTCAG CCTCCCAAGT AGCTGAGACC ACAGGTGTGA GCCACCACTC 420
CTGGGTAATG TTTGTATTTT TTTGTAGAGA TGGGGTTTCA CCGTGCTGCC CAGACTGCTC 480
TCAAACCTCT GGGCTCAAGT GATCCACCTG CTTGACCTC CTAAGTGCT GGAATTACAG 540
GTGTGAGCCA CCGTGCTCAG CCGAGTGCTT TTCGTATGTT TTCTGAGCAC GTGGATTTC 600
ATCTCTCTGC ATTCTCTGTT CATCTCAGCC TGTTTGTGCC ATTGAGATAA ATGACTTTTT 660
CTTGGTAACCT TAGAGTACTT TGTGTATTTA CAGGTTAATC CCTTATCAAT TTATATCAGT 720

```

| | |
|---|------|
| TGCTGCTATC TTTTCTTAGA TTTTCTTTT CATTTTAAAA ATTACATTGT TTCAATGAAC | 780 |
| AGAATTTTAA AGTTTAAACG TAGTCCACTT TGTCACATTT CTTTATGACC GGTGCATTTT | 840 |
| AGGGTCTTGT TTAAGAAATC GTTCTTTATC CTGAGGTCAT AAAGATAGTC TACTGTATTT | 900 |
| TCTTTTAAAG GCTGAAAAGG TGTTTTATAT TTAATTTATT TGGGATTGGC TTTTGTGTGG | 960 |
| TGGGGATAAG GATCACAATT TTATTTTATT TTTTTCAC TTGGTTATGC CAGTGGCCCC | 1020 |
| ATTTCCATTT TTTGAATAGT CTTTCTGTGC AGAAAAGACT TCACTAGCAG AGAAGTCCTG | 1080 |
| AGACTTACCC TTCAAAGGC CCCATTACAC AGGCTAGCAC TTGGCGTGCA TCTGAGAACC | 1140 |
| TGGATTTTGG GGTGGTTCCT ATAATGTGGT GTATGCTGAA CACCCACCTT TCCTTCTGGG | 1200 |
| AGTCTGGAAT TTGGGTATAT GTTGGACAGA GGCTGCCTAA GTGACCAGCT TCAACAACAG | 1260 |
| CCCTGGGTGC TGGGTCACTC ATGACCCATA GACAAA ATG CCA CAC ATG TTG TCA | 1314 |
| Met Pro His Met Leu Ser | |
| 1 5 | |
| CAG CTT ATT GCT GGA GGA GTT AGC ACA TCC TGT GTG ACT GCA CTG GGA | 1362 |
| Gln Leu Ile Ala Gly Gly Val Ser Thr Ser Cys Val Thr Ala Leu Gly | |
| 10 15 20 | |
| GAG GAA ACT GGT GCC TGG TTC CCT GTG TAT TTG TCC CAC GCC TCC AGT | 1410 |
| Glu Glu Thr Gly Ala Trp Phe Pro Val Tyr Leu Ser His Ala Ser Ser | |
| 25 30 35 | |
| CCC TTT GCT GAT CTC GTT TTT TGT CCT TTT GCT GAG ATA AAT CAC AGC | 1458 |
| Pro Phe Ala Asp Leu Val Phe Cys Pro Phe Ala Glu Ile Asn His Ser | |
| 40 45 50 | |
| CAG GAG TAT GAC AAT ATG CGG GGT CCT GTG AGT CCT CCT AAC AAA CAG | 1506 |
| Gln Glu Tyr Asp Asn Met Arg Gly Pro Val Ser Pro Pro Asn Lys Gln | |
| 55 60 65 70 | |
| TTC AAT CTG GGG GTG ATC TTT GGG ATC CCC AAC AAC TGT CGT TTC CCC | 1554 |
| Phe Asn Leu Gly Val Ile Phe Gly Ile Pro Asn Asn Cys Arg Phe Pro | |
| 75 80 85 | |
| ACT GAT AAT AAA ATA ACT GAG AAG CAG CTA TTG GGC AAT GTT CTG AAC | 1602 |
| Thr Asp Asn Lys Ile Thr Glu Lys Gln Leu Leu Gly Asn Val Leu Asn | |
| 90 95 100 | |
| TAC CCT TGAACATTCA TGTCTTCATC TGAACATCCA TCTACTACCC CTGATTTTTT | 1658 |
| Tyr Pro | |
| 104 | |
| CAGTGCAGGG TGCATATCCT GTATCACCCA ATAAATGGTC ATTGATCACC ATAGGAAAGG | 1718 |
| AACAGTGAAA GCTCCACGGT GGTTTGGAGG AAGGTGGCAG GCATTGAGCG GTAACTTTTT | 1778 |
| TGAGCAGATA GATTTTATGT TTTTGCAATG AGTGAAATAA ATTTTCCCAT ATCTATTTAA | 1838 |
| GGTTGGCAAT CATTATCTTT TTATCATCTT GGAACATTTG GAATTCCTTT AATATGTTTA | 1898 |
| GTTAGGAATT TTCTACCTTC CTCATCTTGT CCGATAGTTT AAAATCCAC AGTTATTTCA | 1958 |
| CGGGCTCCTC ATACCTGCCT GTGTGATTTC TAACATGTCA CGCTATGCAA CCAGTTGCTT | 2018 |
| TTACTTGTAG AGTGTTCCTT TAGGTAATAG CTTATTATTG GTTATGTGAT TACAGTGTGT | 2078 |
| TAAAGACAGG TCTGTAGTTA TGTAAATGC CGTTTCTCTG AGTATCATGG TCATTTCAC | 2138 |
| ATATTTCTCT ATTCATGTAT TTGTAAGAAT ATATCTATTT TTGCAGTATT TTATTTATTT | 2198 |
| ATTTTATTTT ATTTTCTGAA ACGGAGCCTT GTTCTGTAC CTAGGCTGGA GTGCAGTGGT | 2258 |
| GTGATCTCGA CTCAGTGTGA CCTCCCCCTC CCAGGTTCAA GCGATTCTCC CGCCTCATCC | 2318 |

bioRxiv preprint doi: <https://doi.org/10.1101/259000>; this version posted September 1, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

```
TCCCAAGTCA TTGGGATTAC AGTCACGTGC CATGAAGCCC TGCTAATTTT TTGTATTTT 2378
AGTAGAGACA GGATTTTACC ATGTTGGCGA TGCTGGTTTC GAACTCCTGG TTTCGAACTC 2438
CTGACCTCAA GTGATCCACC TGCCTCGGCC TCCCAAAGAA CTGGGATTAT GGGCGTGAAC 2498
CACCACGCCA GGTCAGTTT GCAGTGTTT AAATACTGTT GTCTTTGAGA GGAGAGAGGC 2558
ACGCACATAG ACTATGGTGA TTACCATCAT ATACTGGAAA GTGCAAAGTG TAGCGCAGTT 2618
AACTGTGAGC CATCTCATCA AACCTAACA GATGTCTCAT TTGTCCATAA AGGGGCTTCT 2678
GTCCCATAGA AATTCATGTA CCCAACCTAC TCTCAACCA TGATTTTCT CTGATGGCCT 2738
GTGTGAACAG ATTAATGGTG TCCATCTAAT TCCTTCCCA CTGGGGGAAA GCAATCATC 2798
AGGCCATTG CAAAACTGC TCTTGGTTGA GCTTCCTGCC TTAATCATA CCCACAGTA 2858
ATGGCGTCCC TTTATCACCG CTAATGACTC TGACATCTCT CTCCACTCAC ATGTGAGCCT 2918
CCTCAGCTCT CGATAACAA GTCTGTCTCG GTTCATTAT TCTACAAAAA AAAAAAAAAA 2978
AAA 2981
```

SEQ ID NO:4
SEQUENCE LENGTH: 1461
SEQUENCE TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: cDNA
ORIGINAL SOURCE
ORGANISM: human
CELL TYPE: leukocyte
SEQUENCE DESCRIPTION: SEQ ID NO:4

```
AATTCGGCAC GAGCAGCTTT CTAGTTGGAT TAGGCAACAG AATCCTTTGA AAATGTGTGT 60
GCACAGACCA GGTGGCTCTC TGGCCAGTG TACTCTGAAA GATGTGTGTC CTGGCCTAGC 120
TGGTTGAGGA AAAGCAGGGC AAGCCTAGCC AAATCACACA TCTTGAACAG CCCTCATTGG 180
TTATACTAAC TTTCCACCT TCTGGTGTGT ATAGGAGATA AAGATGGCAG ACGTGCTATT 240
AGGCTGCCAA TGGGAGTGGG CTCTGATATG GTCTTTCAAA T ATG AAT CAC CCC TGG 296
Met Asn His Pro Trp
1 5
CAT GTG TGT TTC CTG TTT AAG GTT CTC AGG TAT TAC CCA ACT GCA CCA 344
His Val Cys Phe Leu Phe Lys Val Leu Arg Tyr Tyr Pro Thr Ala Pro
10 15 20
ATA TTA AAA TGG ACA CAT ACC GTG TCA TGC AGT TGG TGC CGA AGT GTT 392
Ile Leu Lys Trp Thr His Thr Val Ser Cys Ser Trp Cys Arg Ser Val
25 30 35
TTA AGG GAA GTT GTA GGC AAT GTG AGT TTA TCA GAA AAC TTC ACC ATA 440
Leu Arg Glu Val Val Gly Asn Val Ser Leu Ser Glu Asn Phe Thr Ile
40 45 50
TCA GCA TTT TGC CCT GAG CTT ACA CCA TTC CCA GAT CAA GGT ACA AGC 488
Ser Ala Phe Cys Pro Glu Leu Thr Pro Phe Pro Asp Gln Gly Thr Ser
55 60 65
ACA ATG ATT TCC TTT CTT GAA AAG TTC AAC AAA AGC AAG AGA GAG AGA 536
Thr Met Ile Ser Phe Leu Glu Lys Phe Asn Lys Ser Lys Arg Glu Arg
```

00050672.060490

| | | | | |
|---|-----|-----|-----|------|
| 70 | 75 | 80 | 85 | |
| TTG GAG TTG ATG CTG CAT TTT TAT TCT GTG TTA AGT CTT GAA CCT GCT | | | | 584 |
| Leu Glu Leu Met Leu His Phe Tyr Ser Val Leu Ser Leu Glu Pro Ala | | | | |
| | 90 | 95 | 100 | |
| GTT GCT GAA CAT TGG TCA GGG GAA TTT GAG AAG TGG AAA GTG GGC TTT | | | | 632 |
| Val Ala Glu His Trp Ser Gly Glu Phe Glu Lys Trp Lys Val Gly Phe | | | | |
| | 105 | 110 | 115 | |
| TTT CAC CCT TTG AAA AGA GAG GAT GGA TTC TTC ACC AGA ACT GAC ATT | | | | 680 |
| Phe His Pro Leu Lys Arg Glu Asp Gly Phe Phe Thr Arg Thr Asp Ile | | | | |
| | 120 | 125 | 130 | |
| TAAAAAAGT CAGCGTGGCA CGTTTTAGTA TGTGTGGCAG ATCTAAASAG ACAATATTTT | | | | 740 |
| GATCTCAGGA GTGTTTATTC TTGAACCATC TTCAGAAGTC TAAGATTGGA GAAATAATAA | | | | 800 |
| AATATTGACC ATCCTTCAAA GAGAAAAACA CAGGGCGATC TTTGGCATAG CCTGTCATTT | | | | 860 |
| TGCTCACATT TCACTTCTCT CTCTCCAAC TCAGAGCCCC TGCTGTGGAA CAGGTGCTGT | | | | 920 |
| GCTGGGTGGC AGGGGAGGTC TCTGGCTTTT TTTTTTTTGT ATCTCCGTCT TAACATCTAG | | | | 980 |
| CCTACTGGAG GAAGTGTATT TAATCATCCA CTTATCTGTT AACAATTATC TCTGAGGGCC | | | | 1040 |
| CGTCACATTC AGAGAAGATT CTAGGTTCCT TACAAGTATC CTCTCACTGT GTACATACTA | | | | 1100 |
| AATCAACATC CTGCTGGATT TCCCCAGAC ATCTCCCTTC ATCACCATTG GAGAGTATCC | | | | 1160 |
| TCTAATTGCC AGCCCTATTC ACCATACTCA TCTCATTGTA TCTGGAGTTT TCTGAGAGTG | | | | 1220 |
| ACCGGGGGTG GGATGGACAG GATAATTTAG CAAGAGTGTA TAAGTAAAT CTATATAATA | | | | 1280 |
| AAAGTTATCT CCTGTGCCC CCCATGATCT ATTCTTTATG TAGCAGTCTG AATGAGATTT | | | | 1340 |
| TCAGAAACAA GAACCACTTT ACCTTAGTCT CTCTCTCTTC TTCTTCTTCT TTTCTTTTCT | | | | 1400 |
| TTTTTTTATG TATTATGGC AACAGAGCAA GACCCAGTCT CAGGAAAAAA AAAAAAAAAA | | | | 1460 |
| A | | | | 1461 |

SEQ ID NO:5
SEQUENCE LENGTH: 3329
SEQUENCE TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: cDNA
ORIGINAL SOURCE
ORGANISM: human
CELL TYPE: leukocyte
SEQUENCE DESCRIPTION: SEQ ID NO:5

| | |
|---|-----|
| CCAAAGTGCT GGGATTATAG GCATGAGCCA CTGCGCCCGG CCAGAATACC CTATCCTTAA | 60 |
| ACATGAATTT AGGGGAGGGG AGGACACAAT TCAATCTATA ACAACTATCA CTGGCTGATT | 120 |
| TTGGCAGAGG CCTGTGGCCT CCAGTATTTT GAGGGAGCTG AGGGCCACTG ATCTCTCCAT | 180 |
| ATGCTCTCAA CATCATGGGA CTAGTAGGAT GAAAGCAAGC CTCAGACCAG ATTCTACCTC | 240 |
| AAGCAGGCAC ACAAACATTC ATGCAGCTTC TACTTGGAGC CTGATGAAGT TCAAATTGTT | 300 |
| TGTCCTCTGA GGCTCTCTTT GCATGGAAAT TTCTCCCATG ACAGATGAGA AAGTTCTGGG | 360 |
| GCAGCATTCA GCTTCTAGT TGGATTAGGC AACAGAATCC TTTGAAAATG TCTGTGCACA | 420 |
| GACCAGGTGG CTCTCTGGGC CAGTGTACTC TGAAGATGT GTGTCCTGGC CTAGCTGGTT | 480 |
| GAGGAAAAGC AGGGCAAGCC TAGCCAAATC ACACATCTTG AACAGCCCTC ATTCGTTATA | 540 |

| | | | | | | |
|-------------|---------------|-------------|-------------------------|-------------|-------------|------|
| CTAACCTTCC | CACCCCTCGG | TGTGTATAGG | AGATAAAGAT | GGCAGACGTG | CTATTAGGCT | 600 |
| GCCAATGGGA | GTGGGCTCTG | ATATGGTCTT | TCAAAT ATG | AAT CAC CCC | TGG CAT | 554 |
| | | | Met Asn His Pro Trp His | | | |
| | | | 1 | 5 | | |
| GTG TGT TTC | CTG TTT AAG | GTT CTC AGG | TAT TAC CCA | ACT GCA CCA | ATA | 702 |
| Val Cys Phe | Leu Phe Lys | Val Leu Arg | Tyr Tyr Pro | Thr Ala Pro | Ile | |
| | 10 | 15 | 20 | | | |
| TTA AAA TGG | ACA CAT ACC | GTG TCA TGC | AGT TGG TGC | CGA AGT GTT | TTA | 750 |
| Leu Lys Trp | Thr His Thr | Val Ser Cys | Ser Trp Cys | Arg Ser Val | Leu | |
| | 25 | 30 | 35 | | | |
| AGG GAA GTT | GTA GGC AAT | GTG AGT TTA | TCA GAA AAC | TTC ACC ATA | TCA | 798 |
| Arg Glu Val | Val Gly Asn | Val Ser Leu | Ser Glu Asn | Phe Thr Ile | Ser | |
| | 40 | 45 | 50 | | | |
| GCA TTT TGC | CCT GAG CTT | ACA CCA TTC | CCA GAT CAA | GGT ACA AGC | ACA | 846 |
| Ala Phe Cys | Pro Glu Leu | Thr Pro Phe | Pro Asp Gln | Gly Thr Ser | Thr | |
| | 55 | 60 | 65 | 70 | | |
| ATG ATT TCC | TTT CTT GAA | AAG TTC AAC | AAA AGC AAG | AGA GAG AGA | TTG | 894 |
| Met Ile Ser | Phe Leu Glu | Lys Phe Asn | Lys Ser Lys | Arg Glu Arg | Leu | |
| | 75 | 80 | 85 | | | |
| GAG TTG ATG | CTG CAT TTT | TAT TCT GTG | TTA AGT CTT | GAA CCT GCT | TTT | 942 |
| Glu Leu Met | Leu His Phe | Tyr Ser Val | Leu Ser Leu | Glu Pro Ala | Phe | |
| | 90 | 95 | 100 | | | |
| GCT GAA CAT | TGG TCA GGG | GAA TTT GAG | AAG TGG AAA | GTG GGC TTT | TTT | 990 |
| Ala Glu His | Trp Ser Gly | Glu Phe Glu | Lys Trp Lys | Val Gly Phe | Phe | |
| | 105 | 110 | 115 | | | |
| CAC CCT TTG | AAA AGA GAG | GAT GGA TTC | TTC ACC AGA | ACT GAC ATT | TAAAAA | 1041 |
| His Pro Leu | Lys Arg Glu | Asp Gly Phe | Phe Thr Arg | Thr Asp Ile | | |
| | 120 | 125 | 130 | | | |
| AAGTACCGT | GGCACGTTT | AGTATGTGTG | GCAGATCTAA | AGAGACAATA | TTTTGATCTC | 1101 |
| AGGAGTGT | TTT ATTCTGAAC | CATTTTCAGA | ACTCTAAGAT | TTGAGAAATA | ATAAAAATATT | 1161 |
| GACCATCCTT | CAAAAGAGAA | AACACAGGGC | GATCTTTGGC | ATAGCCCTGC | ATTTTGTCTCA | 1221 |
| CATTTCACCT | CTCTCTCTCC | AACTTCAGAG | CCCGCTGCTG | GGAACAGGTG | CTGTGCTGGG | 1281 |
| TGGCAGGGGA | GGTCTCTGGC | TTTTTTTTTT | TGATCTCCGT | CTTAACATCT | AGCCTACTGG | 1341 |
| AGGAAGTGTA | TTTAATCATC | CACCTTATCTG | TTAACAATTA | TCTCTGAGGG | CCCGTCACAT | 1401 |
| TCACAGAGA | TTCTAGGTTC | TCTACAGTA | TCCTCTCACT | GTGTACATAC | TAAATCAACA | 1461 |
| TCCTGTGGA | TTTCCCCCAG | ACATCTCCCT | TCATCACCAT | TGGAGAGTAT | CCTCTAATTG | 1521 |
| CCAGCCCAT | TCACCACAT | CATCTCATTT | GATCTGGAGT | TTTCTGAGAG | TGACCGGGGG | 1581 |
| TGGGATGGAC | AGGATAATT | AGCAAGAGTG | TATAAGTAAA | ATCTATATAA | TAAAAGTTAT | 1641 |
| CTCCCTGTGC | CCCCCATGAT | CTATTCTTTA | TGTAGCAGTC | TGAATGAGAT | TTTCAGAAAC | 1701 |
| AAGAACCACT | TTACCTTAGT | CTCTCTCTCT | TCTCTCTCT | CTTTCTTTT | CTTTTTTTTT | 1761 |
| AGTATTATGG | GGATCTGTTT | CTGTGTCCTA | GGGTGGAGTG | CAGTGGTATG | ATCTTGGCTC | 1821 |
| ACAGCAGCGT | TGAACCTCCC | GGCTCAAGTG | GTCTCTCTGC | CTCTGCTTCC | CTAGTAGCTA | 1881 |
| GGACATGCAGT | TTTGTGCCAC | CACACCTGGC | TAATTGAAAA | AGAAATTTTT | TTTTCACATG | 1941 |
| AGACAGGTGC | TTGTCTATGC | CCCCAGCTGG | TCTCAAACTC | CTGGGCTCAA | GTGATCTCTC | 2001 |

| | | | | | | |
|-------------|-------------|-------------|-------------|-------------|-------------|------|
| TGTCATACCC | TCCCAAAGTG | TTGGAATTAC | AGGTTGTGAGC | TACTATACTC | GGCCAGTACC | 2061 |
| CTTCTCAAAA | CACCTTCAGCA | CTTCCCATTG | CACCTTGGGTT | GAAATTCOCA | CCACTCACTG | 2121 |
| GGGCCCCACAA | GACTCTTCAA | GACTGAAATCC | TGCTCAACA | TTGTGACCTG | CCCCTACCA | 2181 |
| CCTGCAGCCT | CACCTTGCTGT | GCTCCAGCCA | TGTGGATCTT | CCTCCTGTCT | CTAAAGACTGC | 2241 |
| CTCAGGTGCAT | TGCAACCTGC | TGTCTTCCG | AAGGCTGTG | TGATTTGCCAT | CAGTCAGTCT | 2301 |
| TAGCTCGTAT | AGCTCTCTGG | AGACACCTCT | TGACGCCAAC | GACCTCAAAAG | AATCTCTCT | 2361 |
| TATCATGTCA | CTCTGTTTTA | TTTATTIATT | TAGAGATGGA | GTCCTGCTCT | GTCAACCAGG | 2421 |
| CTGGAGTGCA | GTCGGCGCAT | CTCTGCTCAC | TGCAAGCTCC | ACCTCCTGGG | TTCATGCCGT | 2481 |
| TCTCCTGCCT | CAGCCTCCTG | AGTAACTGGG | ACTATGGGCA | CCCACCACCTA | CACCCGGGCTA | 2541 |
| ATTTTTTGTGA | TTTTTGTAGTG | GGATGGGGTT | TCACCTGTGT | AGCCAGGATG | GTCTTGATCT | 2601 |
| CCTGACCTTG | TGATCTGCCCT | GCTCCACCT | CCCAAAGTGT | TTTATTIATT | TTAAAGGCCAT | 2661 |
| GTATCACTG | CTGAAAAATTA | GCTCTTCTCT | TCTTTTCTCT | TGTTATCATC | CATTGCCCGG | 2721 |
| AACCAGAATA | GAAGTTCCCTG | AGGCCAGAAC | TTCTGTCTCT | CTGCCCCCTCA | CTATGTGTCT | 2781 |
| CTGGACATATA | CCCCAGTGGC | TGCTGCTCT | AAAGTAAAAAT | CTTAGTAAAT | ATTACTGTTG | 2841 |
| ACTAAATAAA | TGAATAAACT | CCTTTTAATG | CCCTTTTGGG | AGTTGCCAAG | TAAAGAAATG | 2901 |
| GATCCCTTTT | TAAGATTACA | CTTTTGGCTA | TTGATCTGTG | TGTCTGGAAC | AAGATACAGT | 2961 |
| TGTAAGATAC | TACGATTGGGA | CATGACATAA | GTTGAGCTGT | TTAGCGTTTT | AGTAATAAGA | 3021 |
| ATCCAGGATG | TGTCGGGGTG | CGGTGCTCAC | GCCTGTAATC | CTAGCATTTT | GGGAGACCGA | 3081 |
| GGCGGGCAGA | TCACGAGGTC | AGCAGTTTGA | GACCAAGCTG | ACCAACATGG | TGAAACCCCG | 3141 |
| TCTCTACTAA | AAAATACAGA | AATTAGCCGG | GTGTGTTGGT | GTCACACTGT | AGTCCTAGCT | 3201 |
| ACTCAGGAGG | CTGGGGCAGG | AGAATTTCTT | GAACCCGGGA | GGCGGAGGTT | GCAGTGAGCC | 3261 |
| GAGATCACAC | CAGTGCACTC | CAGCCTGGGC | AACAGAGCAA | GACCCAGTCT | CAGGAAAAAA | 3321 |
| AAAAAAA | | | | | | 3321 |

```

SEQ ID NO:6
SEQUENCE LENGTH: 2276
SEQUENCE TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: cDNA
ORIGINAL SOURCE
ORGANISM: human
CELL TYPE: leukocyte
SEQUENCE DESCRIPTION: SEQ ID NO:6
CTGAAC TGGG AGTCAGGTGG TTGACTTGTG CCTGGCTGCA GTAGCAGGG CATCTCCCTT 60
GCACAGTTCT CCTCCTCGGC CTGCCCAAGA GTCCACCAGG CC ATG GAC GCA GTG 114
                                         Met Asp Ala Val
                                         1
GCT GTG TAT CAT GGC AAA ATC AGC AGG GAA ACC GGC GAG AAG CTC CTG 162
Ala Val Tyr His Gly Lys Ile Ser Arg Glu Thr Gly Glu Lys Leu Leu
    5             10             15             20
CTT GCC ACT GGG CTG GAT GGC AGC TAT TTG CTG AGG GAC AGC GAG AGC 210
Leu Ala Thr Gly Leu Asp Gly Ser Tyr Leu Arg Asp Ser Glu Ser
          25             30             35

```

| | |
|--|------|
| GTG CCA GGC GTG TAC TGC CTA TGT GTG CTG TAT CAC GGT TAC ATT TAT | 258 |
| Val Pro Gly Val Tyr Cys Leu Cys Val Leu Tyr His Gly Tyr Ile Tyr | |
| 40 45 50 | |
| ACA TAC CGA GTG TCC CAG ACA GAA ACA GGT TCT TGG AGT GCT GAG ACA | 306 |
| Thr Tyr Arg Val Ser Gln Thr Glu Thr Gly Ser Trp Ser Ala Glu Thr | |
| 55 60 65 | |
| GCA CCT GGG GTA CAT AAA AGA TAT TTC CGG AAA ATA AAA AAT CTC ATT | 354 |
| Ala Pro Gly Val His Lys Arg Tyr Phe Arg Lys Ile Lys Asn Leu Ile | |
| 70 75 80 | |
| TCA GCA TTT CAG AAG CCA GAT CAA GGC ATT GTA ATA CCT CTG CAG TAT | 402 |
| Ser Ala Phe Gln Lys Pro Asp Gln Gly Ile Val Ile Pro Leu Gln Tyr | |
| 85 90 95 100 | |
| CCA GTT GAG AAG AAG TCC TCA GCT AGA AGT ACA CAA GGT ACT ACA GGG | 450 |
| Pro Val Glu Lys Lys Ser Ser Ala Arg Ser Thr Gln Gly Thr Thr Gly | |
| 105 110 115 | |
| ATA AGA GAA GAT CCT GAT GTC TGC CTG AAA GCC CCA TGAAGAAAAA | 496 |
| Ile Arg Glu Asp Pro Asp Val Cys Leu Lys Ala Pro | |
| 120 125 | |
| TAAACACCT TGTACTTTAT TTCTATAAT TTAATATAT GCTAAGTCTT ATATATTGTA | 556 |
| GATAATACAG TTCGGTGAGC TACAAATGCA TTCTAAAGC CATTGTAGTC CTGTAATGGA | 616 |
| AGCATCTAGC ATGTGCTCAA AGCTGAAATG GACTTTTGTA CATAGTGAGG AGCTTTGAAA | 676 |
| CGAGGATTGG GAAAAGTAAT TCCGTAGGTT ATTTTCAGTT ATTATATTTA CAAATGGGAA | 736 |
| ACAAAAGGAT AATGAATACT TTATAAAGGA TTAATGTCAA TTCTTGCCAA ATATAATAA | 796 |
| AAATAATCCT CAGTTTTTGT GAAAAGCTCC ATTTTATAGTG AAATATTATT TTATAGCTAC | 856 |
| TAATTTTAAA ATGTCCTTGT TGATTGTATG GTGGGAAGTT GGCTGGGTGC CCTTGCTTTT | 916 |
| GCCAAGTTCT CCACTAGCTA TGGTGTCTA GGTCTTTTG GGATTTTGA AGCTGTATAC | 976 |
| TGTGTGCTAA AACAAGCACT AAACAAAGAG TGAAGGATTT ATGTTTAATT CTGAAAGCAA | 1036 |
| CCTTCTTGCC TAGTGTCTG ATATTGGACA GTAAATCCA CAGACCAACC TGGAGTTGAA | 1096 |
| AATCTTATAA TTTAAATAT GCTCTAAACA TGTTTATCGT ATTTGATGCT ACAGGATTTG | 1156 |
| AAATTGTATT ACAAATCCAA TGAATGAGT TTTTCTTTTC ATTTACCTCT GCCCCAGTTG | 1216 |
| TTTCTACTAC ATGGAAGACC TCATTTTGAA GGGAAATTC AGCAGCTGCA CCTCATGAGT | 1276 |
| AACTGATTTG TAACAAGCCT CCTTTTAAAG TAACCCTACA AAACCACTGG AAAGTTTATG | 1336 |
| GTTGTATTAT TTTTAAAAA AATTCCAAGT GATTGAACT TACACGAGAT ACAGAATTTT | 1396 |
| ATGCCGCATT TTCTCTCAC ATTTATATTT TTGTGATTTT GTGATTGATT ATATGTCACT | 1456 |
| TTGCTACAGG GCTCACAGAA TTCATTCCT CAACAAACAT AATAGGGCGC TGAGGGCATA | 1516 |
| GAAGTAAAAA CACCTGGTCC CTGCTCTCAG TTCACTGTCT TGTTGGACGA GAAAACAATA | 1576 |
| ACGATAAAG ACAGTGAAAG AAAATAACGA TAAAAGACAG TGAAGAAAAA TAACAATAAA | 1636 |
| AGACAAGGAA AAAATAACAA TGAAGTTGA TAAGTACATG ATAAGCGAGG TTCCCCGTGT | 1696 |
| GTAGGTAGAT CTGGTCTTTA GAGGCAGATA GATAGGTCAG TGCAAATACT CTGGTCCATG | 1756 |
| GGCCATATGA AAAGGCTAAG CTTCCTGTA AAATAATAAC TGGGAATTCT GGGTTGTGTA | 1816 |
| TGGGTGTTGG TGAACCTGGT TTTAATTAGT GAACTGCTGA GAGACAGAGC TATTCTCCAT | 1876 |
| GTA CTGGCAA GACCTGATT CTGAGCATT AATATGGATG CCGTGGGAGT ACAAAGTGG | 1936 |
| AGTGTGCCT GAGTAATGCA TTATGGGTGG TTACCATTT CTTGAGGTAA AAGCATCACA | 1996 |
| TGAAC TTGTA AAGGAATTGA AAAATCCTAC TTTCATAATA AGTTGCATAG GTTTAATAAT | 2056 |

09090572.060490

TTTAAATTAT ATGGCTTGAG TTAAATTGT AATAGGCGTA ACTAATTTTA ACTCTATAAT 2116
GTGTTTCATTC TGGAAATAATC CTAACATAT GAATTATGTT TGCATGTTCA CTCCAAGAG 2176
CCTTTTTTTG AAAAAAGCT TTTTITGAAT CATCAAGTCT TTCACATTTA AATAAAGTGT 2236
TTGAAAGCTT TATTTAAAAA AAAAAAAAAA AAAAAAAAAA 2276

SEQ ID NO:7
SEQUENCE LENGTH: 165
SEQUENCE TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: cDNA
ORIGINAL SOURCE
ORGANISM: human
CELL TYPE: leukocyte
SEQUENCE DESCRIPTION: SEQ ID NO:7
CACTTATAAA ATGTTAGGGC TTAATATTAT TCATAGATCG AGGATAGTTT CATTCTTAGT 60
CGCCTCCTTA GTCACCTTC CTATACCAAT CTGAGACCAT TTTACAATTT AGAAAAGACA 120
AATAACTGGT TGGTTACTT GATAGTATAA TAACC 155

SEQ ID NO:8
SEQUENCE LENGTH: 278
SEQUENCE TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: cDNA
ORIGINAL SOURCE
ORGANISM: human
CELL TYPE: leukocyte
SEQUENCE DESCRIPTION: SEQ ID NO:8
GAAGGAGAAT ATGAAGAGGT TAGAAAAGNT CNGGNTTCTG TTGGTGAAAT GAAGGATCAA 60
GGGGAAGAGA CATTAAATTA TCCTGATACT ACCATTGACT TGTCTCACCT TCAACCCCAA 120
AGGTCCATCC AGAAATTGGC TTCAAAAGAG GAATCTTCTA ATTCTAGTGA CAGTAAATCA 180
CAGAGCCGGA GACATTTGTC AGCCAAGGAA AGAAGGGAAA TGAAAAGAA AAAACTTCCA 240
AGTGACTCAG GAGATTTAGA AGCGTTAGAG GGAAAGGA 278

SEQ ID NO:9
SEQUENCE LENGTH: 135
SEQUENCE TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: cDNA
ORIGINAL SOURCE
ORGANISM: human
CELL TYPE: leukocyte

09090572-050495

SEQUENCE DESCRIPTION: SEQ ID NO:9
TTCTGACAAT GAGTAAGAAG AAAGAGGGTC TTGCCCTTG GTTATTAAGA TTTATCATAG 60
AGCAATAATA ASTAAATCGG TGTATACCA GCACAGAGAT TAGACAAATA AACCAAGGGA 120
CTGGACTAAA TAAGC 135

SEQ ID NO:10
SEQUENCE LENGTH: 197
SEQUENCE TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: cDNA
ORIGINAL SOURCE
ORGANISM: human
CELL TYPE: leukocyte
SEQUENCE DESCRIPTION: SEQ ID NO:10
ATGGTACCCA GTTCAAATT AACATGGTTA TTTTACTTGT GTTCCCAAAT TTAACATTAG 60
GGAATTTTGG GTTGTGGGTC TGTATCACT AGAAAAATAT ATATATTGGT GCTGAAGATA 120
ATTTTGAGAT AATTAGACAA GACAGTTTAG CATTTACAAG AACAAAGTTG GCAGTTGAAG 180
AATCTATTTA TATGACT 197

SEQ ID NO:11
SEQUENCE LENGTH: 137
SEQUENCE TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: cDNA
ORIGINAL SOURCE
ORGANISM: human
CELL TYPE: leukocyte
SEQUENCE DESCRIPTION: SEQ ID NO:11
CCACCGCACC TGGCTGATGC TTTTCTATCT GACTTCTTTC AGAGGACCCT GAAAGACACT 60
AAGTGGAAATC TTTCCTTGAA GTCCTCAAG CTAACAACAT TCTCTGGAAA GATCACCTCT 120
GTTCAGTCTT GGTCTCT 137

SEQ ID NO:12
SEQUENCE LENGTH: 274
SEQUENCE TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: cDNA
ORIGINAL SOURCE
ORGANISM: human
CELL TYPE: leukocyte
SEQUENCE DESCRIPTION: SEQ ID NO:12

0000572.050495

CGTTTACAGA TTCTCTTGGG GCTGGCGGTG GAACTACAAA GGGATCGGTG CCTATATCAC 60
AATACCAAAC TTGATAATAA TCTAGATTCT GTGTYTCTGC TTATAGACCA TGTITGTAGT 120
AGGTAAGAGG AAAACTTCCT ATATTCTGAA ACAGCCTAAC ATTTTACAAA ATTTTAGTTT 180
TCTTTTTTAG AGTCTTATCC TGTAGCTATA TAACAGTTCA TGTCTGATTT AGCATTGTGT 240
CACGAGTAAA GCTGGAAC TAAGAAATTGA AAAT 274

SEQ ID NO:13
SEQUENCE LENGTH: 171
SEQUENCE TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: cDNA
ORIGINAL SOURCE
ORGANISM: human
CELL TYPE: leukocyte
SEQUENCE DESCRIPTION: SEQ ID NO:13
GATTAGGTGA CCTTCCTTGA ARAGCCACGG GTTCCCATTA TCGAAATGCT ATTCATTACC 60
CGAGTCACCT ANGTTCTTAC AAAGGAAGCG AGAAAATTGC TTTGTGTTGG CCATGCCCTT 120
TTTGCANAGG TTCCTAAGTA TAGTCGCCAN AATTTTTTTA ATGGCCTAAA G 171

SEQ ID NO:14
SEQUENCE LENGTH: 161
SEQUENCE TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: cDNA
ORIGINAL SOURCE
ORGANISM: human
CELL TYPE: leukocyte
SEQUENCE DESCRIPTION: SEQ ID NO:14
AGGGGCGCTT GTTCTGCTCT CAGCAGATTG GTTACACGGC TCAGGTGGTG GCGATGACTT 60
AATTCCTAGC CCAAGAAGAA TATAATGTTA AAACGGTTA TGTAATTTT GTGCCTCTCC 120
TTTTTAATGC AGTATTTAGT TCAGATGTTG GCGATTTTTC A 161

SEQ ID NO:15
SEQUENCE LENGTH: 323
SEQUENCE TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: cDNA
ORIGINAL SOURCE
ORGANISM: human
CELL TYPE: leukocyte
SEQUENCE DESCRIPTION: SEQ ID NO:15

| SEQUENCE DESCRIPTION: SEQ ID NO:22 | | |
|------------------------------------|------------|------------|
| GTAGCATTTT | GGCAGAACCA | TTGTTAAITA |
| CCAGATTATT | GAGCRGCCCA | ATGAATGCTT |
| TTTTTTTCCA | ATTCTTTGTA | CTATTTTTTT |
| ATAGGGTATT | TGTTGATAAA | TAATTCATCA |
| CAGAGTGGAA | AAATAAAAAA | AGGGGGAGAA |
| TCCTAGTGAC | AAAAGTTGAG | AAACTACCTT |
| | | TAAATAAGAC |
| | | AGTGAGGTAA |
| | | CAAAATGT |

SEQUENCE DESCRIPTION: SEQ ID NO:23

| | | | | | | |
|------------|------------|------------|------------|------------|------------|-----|
| TGGAATAGCC | AGGAGAATTC | TGGAAGAACT | GAATAATGAG | GTAGGGCTTC | CCTTCGTAT | 60 |
| TTTGAAGTGC | AGATTACACT | ATGTAAAAAC | ATTAGGAAC | GGCAGTGAA | TAGACAGATC | 120 |
| AATAGTTAAT | AGCTGTATTA | AGCAGAAAT | GGTGTAGGA | CAACAGGCTA | ACTAACCTG | 180 |
| TCACTTGTA | TGCTAAAAAT | AAGCTAGAT | AGAGTCCTC | | | 219 |

- 85 -

Figure 1 displays 12 line drawings of the dorsal view of the carapace of various shrimp species. The drawings are arranged vertically and labeled with numbers 1 through 12. Each drawing shows the outline and internal structure of the carapace, including the rostrum, eyes, and various setae. The species names are listed to the right of each drawing.

| Figure | Species |
|--------|--------------------|
| 1 | <i>Stomatopoda</i> |
| 2 | <i>Stomatopoda</i> |
| 3 | <i>Stomatopoda</i> |
| 4 | <i>Stomatopoda</i> |
| 5 | <i>Stomatopoda</i> |
| 6 | <i>Stomatopoda</i> |
| 7 | <i>Stomatopoda</i> |
| 8 | <i>Stomatopoda</i> |
| 9 | <i>Stomatopoda</i> |
| 10 | <i>Stomatopoda</i> |
| 11 | <i>Stomatopoda</i> |
| 12 | <i>Stomatopoda</i> |

Figure 1 displays 12 line drawings of the dorsal view of the carapace of various shrimp species. The drawings are arranged vertically and labeled with numbers 1 through 12. Each drawing shows the outline and internal structure of the carapace, including the rostrum, eyes, and various setae. The species names are listed to the right of each drawing.

| Figure | Species |
|--------|--------------------|
| 1 | <i>Stomatopoda</i> |
| 2 | <i>Stomatopoda</i> |
| 3 | <i>Stomatopoda</i> |
| 4 | <i>Stomatopoda</i> |
| 5 | <i>Stomatopoda</i> |
| 6 | <i>Stomatopoda</i> |
| 7 | <i>Stomatopoda</i> |
| 8 | <i>Stomatopoda</i> |
| 9 | <i>Stomatopoda</i> |
| 10 | <i>Stomatopoda</i> |
| 11 | <i>Stomatopoda</i> |
| 12 | <i>Stomatopoda</i> |

ACAATATTTA CTTGCAAAAG AGCCACACAGA GACATGTCAA TGAAGTCATA G

291

SEQ ID NO:33

SEQUENCE LENGTH: 230

SEQUENCE TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

ORIGINAL SOURCE

ORGANISM: human

CELL TYPE: leukocyte

SEQUENCE DESCRIPTION: SEQ ID NO:33

Met Glu Asp Gly Phe Leu Asp Asp Gly Arg Gly Asp Gln Pro Leu His
 1 5 10 15
 Ser Gly Leu Gly Ser Pro His Cys Phe Ser His Gln Asn Gly Glu Arg
 20 25 30
 Val Glu Arg Tyr Ser Arg Lys Val Phe Val Gly Gly Leu Pro Pro Asp
 35 40 45
 Ile Asp Glu Asp Glu Ile Thr Ala Ser Phe Arg Arg Phe Gly Pro Leu
 50 55 60
 Ile Val Asp Trp Pro His Lys Ala Glu Ser Lys Ser Tyr Phe Pro Pro
 65 70 75 80
 Lys Gly Tyr Ala Phe Leu Leu Phe Gln Asp Glu Ser Ser Val Gln Ala
 85 90 95
 Leu Ile Asp Ala Cys Ile Glu Glu Asp Gly Lys Leu Tyr Leu Cys Val
 100 105 110
 Ser Ser Pro Thr Ile Lys Asp Lys Pro Val Gln Ile Arg Pro Trp Asn
 115 120 125
 Leu Ser Asp Ser Asp Phe Val Met Asp Gly Ser Gln Pro Leu Asp Pro
 130 135 140
 Arg Lys Thr Ile Phe Val Gly Gly Val Pro Arg Pro Leu Arg Ala Val
 145 150 155 160
 Glu Leu Ala Met Val Met Asp Arg Leu Tyr Gly Gly Val Cys Tyr Ala
 165 170 175
 Gly Ile Asp Thr Asp Pro Glu Leu Lys Tyr Pro Lys Gly Ala Gly Arg
 180 185 190
 Val Ala Phe Ser Asn Gln Gln Ser Tyr Ile Ala Ala Ile Ser Ala Arg
 195 200 205
 Phe Val Gln Leu Gln His Gly Glu Ile Asp Lys Arg Val Ser Leu Ile
 210 215 220
 Leu His Phe Gly Lys Phe
 225 230

SEQ ID NO:34

SEQUENCE LENGTH: 143

```

SEQUENCE TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
ORIGINAL SOURCE
ORGANISM: human
CELL TYPE: leukocyte
SEQUENCE DESCRIPTION: SEQ ID NO:34
Met Gly Ser Asp Lys Arg Val Ser Arg Thr Glu Arg Ser Gly Arg Tyr
  1             5             10             15
Gly Ser Ile Ile Asp Arg Asp Asp Arg Asp Glu Arg Glu Ser Arg Ser
      20             25             30
Arg Arg Arg Asp Ser Asp Tyr Lys Arg Ser Ser Asp Asp Arg Arg Gly
      35             40             45
Asp Arg Tyr Asp Asp Tyr Arg Asp Tyr Asp Ser Pro Glu Arg Glu Arg
      50             55             60
Glu Arg Arg Asn Ser Asp Arg Ser Glu Asp Gly Tyr His Ser Asp Gly
      65             70             75             80
Asp Tyr Gly Glu His Asp Tyr Arg His Asp Ile Ser Asp Glu Arg Glu
      85             90             95
Ser Lys Thr Ile Met Leu Arg Gly Leu Pro Ile Thr Ile Thr Glu Ser
      100            105            110
Asp Ile Arg Glu Met Met Glu Ser Phe Glu Gly Pro Gln Pro Ala Asp
      115            120            125
Val Arg Leu Met Lys Arg Lys Thr Gly Glu Ser Leu Leu Ser Ser
      130            135            140            143

```

```

SEQ ID NO:35
SEQUENCE LENGTH: 104
SEQUENCE TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
ORIGINAL SOURCE
ORGANISM: human
CELL TYPE: leukocyte
SEQUENCE DESCRIPTION: SEQ ID NO:35
Met Pro His Met Leu Ser Gln Leu Ile Ala Gly Gly Val Ser Thr Ser
  1             5             10             15
Cys Val Thr Ala Leu Gly Glu Glu Thr Gly Ala Trp Phe Pro Val Tyr
          20             25             30
Leu Ser His Ala Ser Ser Pro Phe Ala Asp Leu Val Phe Cys Pro Phe
          35             40             45
Ala Glu Ile Asn His Ser Gln Glu Tyr Asp Asn Met Arg Gly Pro Val
          50             55             60
Ser Pro Pro Asn Lys Gln Phe Asn Leu Gly Val Ile Phe Gly Ile Pro

```

65 70 75 80
 Asn Asn Cys Arg Phe Pro Thr Asp Asn Lys Ile Thr Glu Lys Gln Leu
 85 90 95
 Leu Gly Asn Val Leu Asn Tyr Pro
 100

SEQ ID NO:36
 SEQUENCE LENGTH: 133
 SEQUENCE TYPE: amino acid
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 ORIGINAL SOURCE
 ORGANISM: human
 CELL TYPE: leukocyte
 SEQUENCE DESCRIPTION: SEQ ID NO:36
 Met Asn His Pro Trp His Val Cys Phe Leu Phe Lys Val Leu Arg Tyr
 1 5 10 15
 Tyr Pro Thr Ala Pro Ile Leu Lys Trp Thr His Thr Val Ser Cys Ser
 20 25 30
 Trp Cys Arg Ser Val Leu Arg Glu Val Val Gly Asn Val Ser Leu Ser
 35 40 45
 Glu Asn Phe Thr Ile Ser Ala Phe Cys Pro Glu Leu Thr Pro Phe Pro
 50 55 60
 Asp Gln Gly Thr Ser Thr Met Ile Ser Phe Leu Glu Lys Phe Asn Lys
 65 70 75 80
 Ser Lys Arg Glu Arg Leu Glu Leu Met Leu His Phe Tyr Ser Val Leu
 85 90 95
 Ser Leu Glu Pro Ala Val Ala Glu His Trp Ser Gly Glu Phe Glu Lys
 100 105 110
 Trp Lys Val Gly Phe Phe His Pro Leu Lys Arg Glu Asp Gly Phe Phe
 115 120 125
 Thr Arg Thr Asp Ile
 130

SEQ ID NO:37
 SEQUENCE LENGTH: 133
 SEQUENCE TYPE: amino acid
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 ORIGINAL SOURCE
 ORGANISM: human
 CELL TYPE: leukocyte
 SEQUENCE DESCRIPTION: SEQ ID NO:37
 Met Asn His Pro Trp His Val Cys Phe Leu Phe Lys Val Leu Arg Tyr

MOLECULE TYPE: other nucleic acid, synthetic DNA
SEQUENCE DESCRIPTION: SEQ ID NO:45
GTGGATCTGG ATTTTGTCA TATGT

```

SEQ ID NO:46
SEQUENCE LENGTH: 25
SEQUENCE TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: other nucleic acid, synthetic DNA
SEQUENCE DESCRIPTION: SEQ ID NO:46
GTTTGTGATT ATAACCAAC ATGTG

```

SEQ ID NO:47
SEQUENCE LENGTH: 25
SEQUENCE TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: other nucleic acid, synthetic DNA
SEQUENCE DESCRIPTION: SEQ ID NO:47
GAAGGGGAAG AGACATTAA TTATC

SEQ ID NO:48
SEQUENCE LENGTH: 24
SEQUENCE TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: other nucleic acid, synthetic DNA
SEQUENCE DESCRIPTION: SEQ ID NO:48
GCTTCTAAAT CTCCTGAGTC ACTT

SEQ ID NO:49
SEQUENCE LENGTH: 24
SEQUENCE TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: other nucleic acid, synthetic DNA
SEQUENCE DESCRIPTION: SEQ ID NO:49
GACAATGAGT AAGAAGAAAG AGGG

SEQ ID NO:60
SEQUENCE LENGTH: 24
SEQUENCE TYPE: nucleic acid
STRANDEDNESS: single

00000573.050409

TOPOLOGY: linear
MOLECULE TYPE: other nucleic acid, synthetic DNA
SEQUENCE DESCRIPTION: SEQ ID NO:50
GTCCAGTCCC TTGGTTTATT TGTC 24

SEQ ID NO:51
SEQUENCE LENGTH: 25
SEQUENCE TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: other nucleic acid, synthetic DNA
SEQUENCE DESCRIPTION: SEQ ID NO:51
GGTACCCAGT TTCAAATTA CATGG 25

SEQ ID NO:52
SEQUENCE LENGTH: 25
SEQUENCE TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: other nucleic acid, synthetic DNA
SEQUENCE DESCRIPTION: SEQ ID NO:52
GATTCTTCAA CTGCCAAACT TGTC 25

SEQ ID NO:53
SEQUENCE LENGTH: 24
SEQUENCE TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: other nucleic acid, synthetic DNA
SEQUENCE DESCRIPTION: SEQ ID NO:53
GCTGATGCTT TTCTATCTGA CTTC 24

SEQ ID NO:54
SEQUENCE LENGTH: 22
SEQUENCE TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: other nucleic acid, synthetic DNA
SEQUENCE DESCRIPTION: SEQ ID NO:54
GACCAGGACT GAACAGAGGT GA 22

SEQ ID NO:55
SEQUENCE LENGTH: 25
SEQUENCE TYPE: nucleic acid

00000572.050490

SEQUENCE TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: other nucleic acid, synthetic DNA
SEQUENCE DESCRIPTION: SEQ ID NO:60
GCCAACATCT GAACTAAATA CTGC 24

SEQ ID NO:61
SEQUENCE LENGTH: 25
SEQUENCE TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: other nucleic acid, synthetic DNA
SEQUENCE DESCRIPTION: SEQ ID NO:61
GTTCA GTGAA TGTTACCTAG AAACA 25

SEQ ID NO:62
SEQUENCE LENGTH: 24
SEQUENCE TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: other nucleic acid, synthetic DNA
SEQUENCE DESCRIPTION: SEQ ID NO:62
GGAGTGAAAA CTGTCTTGT CATC 24

SEQ ID NO:63
SEQUENCE LENGTH: 25
SEQUENCE TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: other nucleic acid, synthetic DNA
SEQUENCE DESCRIPTION: SEQ ID NO:63
GTATGACAAA TAGTTTCTGC CTGAT 25

SEQ ID NO:64
SEQUENCE LENGTH: 25
SEQUENCE TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: other nucleic acid, synthetic DNA
SEQUENCE DESCRIPTION: SEQ ID NO:64
GATTAACAAA GATGTACAGA CTGAG 25

SEQ ID NO:65

```
SEQUENCE LENGTH: 24
SEQUENCE TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: other nucleic acid, synthetic DNA
SEQUENCE DESCRIPTION: SEQ ID NO:65
GAGACAGCAT TCAGATATAG ACGG
```

```

SEQ ID NO:66
SEQUENCE LENGTH: 22
SEQUENCE TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: other nucleic acid, synthetic DNA
SEQUENCE DESCRIPTION: SEQ ID NO:66
GCGTGGAATC AAATGGAGTG GC

```

SEQ ID NO:67
SEQUENCE LENGTH: 24
SEQUENCE TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: other nucleic acid, synthetic DNA
SEQUENCE DESCRIPTION: SEQ ID NO:67
GATGGCCTGT GTGAACAGAT TAAT

SEQ ID NO:68
SEQUENCE LENGTH: 24
SEQUENCE TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: other nucleic acid, synthetic DNA
SEQUENCE DESCRIPTION: SEQ ID NO:68
GAGAGAGATG TCAGAGTCAT TAGC

SEQ ID NO:69
SEQUENCE LENGTH: 24
SEQUENCE TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: other nucleic acid, synthetic DNA
SEQUENCE DESCRIPTION: SEQ ID NO:69
GATCCCCACA ATTTCGTGTG ATTG

SEQ ID NO:70
SEQUENCE LENGTH: 25
SEQUENCE TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: other nucleic acid, synthetic DNA
SEQUENCE DESCRIPTION: SEQ ID NO:70
GTTCCCTTAA AATAATGTGG TAATG

SEQ ID NO:71
SEQUENCE LENGTH: 23
SEQUENCE TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: other nucleic acid, synthetic DNA
SEQUENCE DESCRIPTION: SEQ ID NO:71
GAGGATACTC TCCAATGGTG ATG

SEQ ID NO:72
SEQUENCE LENGTH: 24
SEQUENCE TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: other nucleic acid, synthetic DNA
SEQUENCE DESCRIPTION: SEQ ID NO:72
GTCTTAACAT CTAGCCTACT GGAG

SEQ ID NO:73
SEQUENCE LENGTH: 24
SEQUENCE TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: other nucleic acid, synthetic DNA
SEQUENCE DESCRIPTION: SEQ ID NO:73
GAGAGGAGCC ATGTATACAA ACCA

SEQ ID NO:74
SEQUENCE LENGTH: 26
SEQUENCE TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: other nucleic acid, synthetic DNA
SEQUENCE DESCRIPTION: SEQ ID NO:74
GCACGCAGGA TCAGATATAG TAATTC


```

SEQ ID NO:75
SEQUENCE LENGTH: 24
SEQUENCE TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: other nucleic acid, synthetic DNA
SEQUENCE DESCRIPTION: SEQ ID NO:75
GCTGAAACCT AAGCTGAAGG AAGG

```

```

SEQ ID NO:76
SEQUENCE LENGTH: 22
SEQUENCE TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: other nucleic acid, synthetic DNA
SEQUENCE DESCRIPTION: SEQ ID NO:76
GTCCTCACC TCAGATCACA CC

```

```

SEQ ID NO:77
SEQUENCE LENGTH: 24
SEQUENCE TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: other nucleic acid, synthetic DNA
SEQUENCE DESCRIPTION: SEQ ID NO:77
GCTATCTACC TGGCAGGAAA AGAG

```

```

SEQ ID NO:78
SEQUENCE LENGTH: 25
SEQUENCE TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: other nucleic acid, synthetic DNA
SEQUENCE DESCRIPTION: SEQ ID NO:78
GAGTTTCTTA CTATGATCTG GATTC

```

SEQ ID NO:79
SEQUENCE LENGTH: 25
SEQUENCE TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: other nucleic acid, synthetic DNA
SEQUENCE DESCRIPTION: SEQ ID NO:79

25

24

26

24

22

- 102 -

1990 1991 1992 1993 1994 1995 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010 2011 2012 2013 2014 2015 2016 2017 2018 2019 2020 2021 2022 2023 2024 2025 2026 2027 2028 2029 2030 2031 2032 2033 2034 2035 2036 2037 2038 2039 2040 2041 2042 2043 2044 2045 2046 2047 2048 2049 2050 2051 2052 2053 2054 2055 2056 2057 2058 2059 2060 2061 2062 2063 2064 2065 2066 2067 2068 2069 2070 2071 2072 2073 2074 2075 2076 2077 2078 2079 2080 2081 2082 2083 2084 2085 2086 2087 2088 2089 2090 2091 2092 2093 2094 2095 2096 2097 2098 2099 2100 2101 2102 2103 2104 2105 2106 2107 2108 2109 2110 2111 2112 2113 2114 2115 2116 2117 2118 2119 2120 2121 2122 2123 2124 2125 2126 2127 2128 2129 2130 2131 2132 2133 2134 2135 2136 2137 2138 2139 2140 2141 2142 2143 2144 2145 2146 2147 2148 2149 2150 2151 2152 2153 2154 2155 2156 2157 2158 2159 2160 2161 2162 2163 2164 2165 2166 2167 2168 2169 2170 2171 2172 2173 2174 2175 2176 2177 2178 2179 2180 2181 2182 2183 2184 2185 2186 2187 2188 2189 2190 2191 2192 2193 2194 2195 2196 2197 2198 2199 2200 2201 2202 2203 2204 2205 2206 2207 2208 2209 2210 2211 2212 2213 2214 2215 2216 2217 2218 2219 2220 2221 2222 2223 2224 2225 2226 2227 2228 2229 2230 2231 2232 2233 2234 2235 2236 2237 2238 2239 2240 2241 2242 2243 2244 2245 2246 2247 2248 2249 2250 2251 2252 2253 2254 2255 2256 2257 2258 2259 2260 2261 2262 2263 2264 2265 2266 2267 2268 2269 2270 2271 2272 2273 2274 2275 2276 2277 2278 2279 2280 2281 2282 2283 2284 2285 2286 2287 2288 2289 2290 2291 2292 2293 2294 2295 2296 2297 2298 2299 2300 2301 2302 2303 2304 2305 2306 2307 2308 2309 2310 2311 2312 2313 2314 2315 2316 2317 2318 2319 2320 2321 2322 2323 2324 2325 2326 2327 2328 2329 2330 2331 2332 2333 2334 2335 2336 2337 2338 2339 2340 2341 2342 2343 2344 2345 2346 2347 2348 2349 2350 2351 2352 2353 2354 2355 2356 2357 2358 2359 2360 2361 2362 2363 2364 2365 2366 2367 2368 2369 2370 2371 2372 2373 2374 2375 2376 2377 2378 2379 2380 2381 2382 2383 2384 2385 2386 2387 2388 2389 2390 2391 2392 2393 2394 2395 2396 2397 2398 2399 2400 2401 2402 2403 2404 2405 2406 2407 2408 2409 2410 2411 2412 2413 2414 2415 2416 2417 2418 2419 2420 2421 2422 2423 2424 2425 2426 2427 2428 2429 2430 2431 2432 2433 2434 2435 2436 2437 2438 2439 2440 2441 2442 2443 2444 2445 2446 2447 2448 2449 2450 2451 2452 2453 2454 2455 2456 2457 2458 2459 2460 2461 2462 2463 2464 2465 2466 2467 2468 2469 2470 2471 2472 2473 2474 2475 2476 2477 2478 2479 2480 2481 2482 2483 2484 2485 2486 2487 2488 2489 2490 2491 2492 2493 2494 2495 2496 2497 2498 2499 2500 2501 2502 2503 2504 2505 2506 2507 2508 2509 2510 2511 2512 2513 2514 2515 2516 2517 2518 2519 2520 2521 2522 2523 2524 2525 2526 2527 2528 2529 2530 2531 2532 2533 2534 2535 2536 2537 2538 2539 2540 2541 2542 2543 2544 2545 2546 2547 2548 2549 2550 2551 2552 2553 2554 2555 2556 2557 2558 2559 2560 2561 2562 2563 2564 2565 2566 2567 2568 2569 2570 2571 2572 2573 2574 2575 2576 2577 2578 2579 2580 2581 2582 2583 2584 2585 2586 2587 2588 2589 2590 2591 2592 2593 2594 2595 2596 2597 2598 2599 2600 2601 2602 2603 2604 2605 2606 2607 2608 2609 2610 2611 2612 2613 2614 2615 2616 2617 2618 2619 2620 2621 2622 2623 2624 2625 2626 2627 2628 2629 2630 2631 2632 2633 2634 2635 2636 2637 2638 2639 2640 2641 2642 2643 2644 2645 2646 2647 2648 2649 2650 2651 2652 2653 2654 2655 2656 2657 2658 2659 2660 2661 2662 2663 2664 2665 2666 2667 2668 2669 2670 2671 2672 2673 2674 2675 2676 2677 2678 2679 2680 2681 2682 2683 2684 2685 2686 2687 2688 2689 2690 2691 2692 2693 2694 2695 2696 2697 2698 2699 2700 2701 2702 2703 2704 2705 2706 2707 2708 2709 2710 2711 2712 2713 2714 2715 2716 2717 2718 2719 2720 2721 2722 2723 2724 2725 2726 2727 2728 2729 2730 2731 2732 2733 2734 2735 2736 2737 2738 2739 2740 2741 2742 2743 2744 2745 2746 2747 2748 2749 2750 2751 2752 2753 2754 2755 2756 2757 2758 2759 2760 2761 2762 2763 2764 2765 2766 2767 2768 2769 2770 2771 2772 2773 2774 2775 2776 2777 2778 2779 2780 2781 2782 2783 2784 2785 2786 2787 2788 2789 2790 2791 2792 2793 2794 2795 2796 2797 2798 2799 2800 2801 2802 2803 2804 2805 2806 2807 2

00000672.060490

SEQUENCE DESCRIPTION: SEQ ID NO:84
GCATAACAAG TGACAGGGTT AGTTA 25

SEQ ID NO:85
SEQUENCE LENGTH: 22
SEQUENCE TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: other nucleic acid, synthetic DNA
SEQUENCE DESCRIPTION: SEQ ID NO:85
GGTGCTOCTT CCTTAACTG GT 22

SEQ ID NO:86
SEQUENCE LENGTH: 23
SEQUENCE TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: other nucleic acid, synthetic DNA
SEQUENCE DESCRIPTION: SEQ ID NO:86
GACTACACAT AAACCCACCC CAG 23

SEQ ID NO:87
SEQUENCE LENGTH: 24
SEQUENCE TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: other nucleic acid, synthetic DNA
SEQUENCE DESCRIPTION: SEQ ID NO:87
GGGTACAGGA TTTCTAAGAA GTGG 24

SEQ ID NO:88
SEQUENCE LENGTH: 25
SEQUENCE TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: other nucleic acid, synthetic DNA
SEQUENCE DESCRIPTION: SEQ ID NO:88
GGAGAAAATT TCAGCTCATC TGAAG 25

SEQ ID NO:89
SEQUENCE LENGTH: 24
SEQUENCE TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear

24

23

24

22

22

- 104 -

Figure 1. The effect of the number of iterations on the accuracy of the proposed algorithm. The accuracy is measured by the percentage of correct classification. The number of iterations is 10, 20, 30, 40, 50, 60, 70, 80, 90, 100, 110, 120, 130, 140, 150, 160, 170, 180, 190, 200, 210, 220, 230, 240, 250, 260, 270, 280, 290, 300, 310, 320, 330, 340, 350, 360, 370, 380, 390, 400, 410, 420, 430, 440, 450, 460, 470, 480, 490, 500, 510, 520, 530, 540, 550, 560, 570, 580, 590, 600, 610, 620, 630, 640, 650, 660, 670, 680, 690, 700, 710, 720, 730, 740, 750, 760, 770, 780, 790, 800, 810, 820, 830, 840, 850, 860, 870, 880, 890, 900, 910, 920, 930, 940, 950, 960, 970, 980, 990, 1000. The accuracy is 0.8, 0.85, 0.9, 0.95, 1.0.

09090573-050498

SEQUENCE LENGTH: 22
SEQUENCE TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: other nucleic acid, synthetic DNA
SEQUENCE DESCRIPTION: SEQ ID NO:109
ATGAAAGTGC CAGTGTGCCA TG 22

SEQ ID NO:110
SEQUENCE LENGTH: 22
SEQUENCE TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: other nucleic acid, synthetic DNA
SEQUENCE DESCRIPTION: SEQ ID NO:110
CCCATCACCA TCTTCCAGGA GC 22

SEQ ID NO:111
SEQUENCE LENGTH: 26
SEQUENCE TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: other nucleic acid, synthetic DNA
SEQUENCE DESCRIPTION: SEQ ID NO:111
TTCACCACCT TCTTGATGTC ATCATA 26